

DOCUMENT RESUME

ED 090 300

TM 003 585

AUTHOR Millman, Jason
TITLE Task and Administrative Facets of Teaching Performance Tests.
PUB DATE 74
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)
EDRS PRICE MF-\$0.75 HC-\$1.50 PLUS POSTAGE
DESCRIPTORS Effective Teaching; *Evaluation Methods; *Performance Tests; *Task Analysis; *Teacher Evaluation; Test Construction; Testing; *Testing Problems
IDENTIFIERS *Teaching Performance Tests

ABSTRACT

The teaching performance test is a recently developed assessment technique designed to sharpen our teacher competence evaluation procedures. It assesses a teacher's ability to promote learner mastery of prespecified instructional objectives during a relatively short lesson designed by the teacher. The principal contributions of this paper are the suggestions of what key facets (i.e., task dimensions and administrative factors) are crucial in the effective use of teaching performance tests. The critical dimensions involved in teaching performance tests fall into two groups-those associated with the nature of the instructional objective and those concerned with the administration of the test. (Author)

Task and Administrative Facets of Teaching Performance Tests¹

Jason Millman

Cornell University

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Teachers are asked to perform many tasks: to dream up objectives and to object to dreaming, to put snow boots on the young child and to put the boot to the older one, to teach concepts of arithmetic and to teach self-concept, to teach a discipline and to discipline, to manage the instruction and to manage to survive, to question and to answer, to test and to be tested, to respect and to be respected, and to go out on strike without striking out.

The performances called teaching are more variable than the winning numbers on a Las Vegas roulette wheel. It is the teacher who goes round and round the rounds--the colors are black and blue, bandits come with two arms, and only bells ring at 7 to 11.

Faced with the multiplicity of teaching functions, the chances of being able to describe meaningfully a teacher's competence by a single number are remote. For people who like to win, long-shots are to be avoided both in the gambling casinos and in the testing enterprises. Taking reasonable risks is rational; inviting the impossible is idiotic.

Teaching performance tests which concentrate on only a part of the teaching act are, in my view, a reasonable but risky strategy for measuring teaching effectiveness. Teaching performance tests, in brief, assess a teacher's ability to promote learner mastery on prespecified instructional objectives during a relatively short lesson designed by the teacher. The rationale behind teaching performance tests is the belief that the single most important function

¹ A paper presented at the Annual Meeting of the American Educational Research Association, April 1974.

ED 090300

585

003

TM

of teaching is to contribute to that intellectual, physical, and emotional growth of pupils which could not be expected to be acquired without the benefit of the teacher's intervention. The use of short lessons, called minilessons, is a practical compromise which makes feasible the measurement of pupil growth in a controlled situation.

Even though only one function of the teaching act is focused upon with teaching performance tests, there is that reasonable risk that measurement of how well a teacher can do even this one task may be beyond our reach. With teaching performance tests, a teacher's effectiveness is measured by the mean performance of the pupils observed at the conclusion of a minilesson. When these average performances are correlated across different attempts by the teachers, unstable and frequently low correlations are found. Thus, at the present state of our technology, our judgment of a teacher's competence changes from trial to trial.

One reason for the low correlations between mean pupil performance scores on successive teaching attempts could be that teaching skill is not generalizable. Teaching performance tests may be perfectly accurate in their assessment of teaching competence, but teaching skill (like happiness) can change from situation to situation. To the extent this is the case, we need to know the dimensions of the teaching situation which most affects a person's teaching skill.

We seek to determine the procedures by which teaching performance tests can be constructed and administered so that they can be an effective tool in the repertoire of today's educational researcher and teacher evaluator. The principal contribution of this paper is to suggest factors which we believe are crucial for the effective use of teaching performance tests. The critical dimensions involved in teaching performance tests appear to fall into two groups, those associated with the nature of the instructional objective and those concerned with the administration of the test.

Task Dimensions

The job of constructing teaching performance tests is similar in many respects to the task facing any test developer. One problem needing resolution is the selection of the specific instructional tasks to be used in the minilessons. The test developer should choose teaching tasks within the prespecified boundaries of content which are sufficiently dissimilar that any underlying variability of a teacher's competence can be elicited. As Humphreys (1962) has stated, "The implication for practice in test construction is deliberately to make the test as heterogeneous as possible within the limits of the definition of what you are trying to measure." (p. 481)

When this advice is applied to the construction of teaching performance tests, it results in the construction of minilessons on which the teacher is most likely to demonstrate variable competence. The choice of instructional task is crucial. For example, if one type of lesson was used exclusively, it might result in Teacher A obtaining student performances superior to that of Teacher B; whereas, had a different type of task been used, Teacher B might be associated with the higher competency assessment. Minilessons serve the same role as items on an achievement test. Just as the achievement test items must be representative of the entire content area to which generalizations about the examinee's status is to be inferred, so must the selection of teaching tasks be representative of the teaching situations about which the competency evaluation is directed.

What, then, are the instructional task dimensions which should be sampled by the minilessons? Two sources for the identification of important task dimensions appear most promising. One source is the nature of the learning expected of the student, and learning theory might be helpful in describing appropriate categories. Gagné (1974), for example, suggests

five learning outcomes: verbal information or knowledge, intellectual skills, cognitive strategies, attitudes, and motor skills. Gagné states that

The five classes are significant because they are distinctive in their characteristics. In particular, this means that instruction, for each of these classes of outcome, has readily distinguishable differences. One does not, for example, design instruction for motor skills to be the same as instruction for attitudes; instruction for information is not the same as instruction for intellectual skills. (p. 4)

If Gagné is correct and different instructional skills are required for each type of learning, then it is quite possible that a teacher competent in teaching for one kind of learning outcome will be less successful with other types. Thus, if generalizations about teaching skill to all kinds of learning are required, then a variety of learning outcomes should be sampled by the minilessons.

A second source for identifying important task dimensions is the research on the relationship between the process variables of teaching and student achievement. Although finding a positive relationship between a particular teacher behavior and student performance does not necessarily mean that an increase in such teacher behavior would cause improved student performance, the likelihood that such is the case is strong enough to warrant attention being given to such teacher behaviors in the design and selection of minilesson tasks.

Let me illustrate. One teacher behavior which has consistently correlated positively with student achievement is ^{an} organized approach to instruction. Using concrete materials is another behavior which, although not having the same degree of empirical support for its relationship with achievement, is nevertheless widely encouraged. One of our presently developed minilessons for young children consists of the objective that the children will be able to provide proper scoring for any given end-of-round layout in a table shuffleboard game. This minilesson, requiring the learning of conditional rules, begs to be taught by an organized,

logical approach and is amenable to the use of concrete materials. Teachers who do not engage in such practices are likely to be less successful than those who do, and the minilesson has promise as a discriminator of teaching ability.

In this discussion of sources for the identification of important task dimensions, no mention has been made of the subject matter of the task, as distinct from the kind of learning required. One might be tempted to argue that by virtue of their greater acquaintanceship with certain subjects and their greater store of experiences on which to draw, some teachers would be advantaged if the minilessons they were to teach consisted of tasks especially familiar to them. It does make sense to limit the situations to be sampled to those dealing with types of subject matter content for which the teacher is expected to be responsible. It should be noted, however, that the meager data there is on the question suggest that familiarity of the teacher with the material is largely unrelated to pupil achievement. (See, e. g., Millman 1974.)

Administrative Facets

It was stated above that the nature of the specific minilessons was a very important factor affecting the assessed competence of a teacher. It was suggested that the measurement of teaching competence should sample instructional tasks broadly consistent with the scope of the intended inferences about teaching skills.

There are, however, a number of other aspects of the teaching situation which influences the effectiveness score of a teacher. Some of these factors are identified below.

1. Student Characteristics. Since effectiveness of a teaching attempt is assessed by the mean score of the students on cognitive or affective tests covering the instructional objectives, it would naturally follow that teachers

who taught learners initially more able ^{or} ~~to~~ having initially more positive attitudes would be at a distinct advantage.

Three ways to reduce this problem are to: (a) randomly assign students to the teacher--thus eliminating any systematic bias in favor of a given teacher, (b) prior to instruction, administer control tests which correlate highly with the dependent variable and adjust the mean class scores to negate possible initial imbalances among the classes, and (c) use tasks containing subject matter which is unfamiliar to the students. The table shuffleboard minilesson cited above is an example of such a task.²

Even if the initial abilities and attitudes of the learners were equated among all the groups being taught, attention still needs to be given to those student characteristics which might interact with the teachers. That is, some teachers might be particularly adept working with a certain type of learner.

Brophy and his colleagues at the University of Texas found, for example, differences in the relationship between teacher behaviors and performance measures for Title I and non-Title I schools. We might speculate that teaching behaviors effective for teaching students of one type of ability or culture or having certain previous instructional experiences may be quite different from the behaviors effective for teaching students differing on these attributes or experiences. Thus, teachers who have a proclivity for a particular teaching

² The decision whether or not to use unfamiliar tasks involves a trade-off. It is true that most of a teacher's instructional time is spent teaching content about which the student already has partial knowledge. Yet, unfamiliar tasks do help to separate what the student already knows from what he is taught. If students can only score at the chance level on the criterion test administered prior to instruction, we can be quite sure that their post-instruction test scores reflect what they have learned as a result of the teacher's efforts.

style might be differentially effective with the various student types. Age particularly may be relevant, for it appears that the nature of learning for an adult is far different from how a young child learns. In any application of teaching performance tests it would be advisable, in the absence of data to the contrary, to employ learner groups similar to those for whom the teacher is expected to have responsibility.

2. Number of Observations. If we think of a minilesson trial as analogous to an item on a test, it becomes clear that we need to sample many teaching opportunities if a reliable estimate of teaching skill is to be obtained. This situation is not too different from research using classroom observation schedules in which ten or more observation periods are viewed by some as needed to secure a stable estimate of classroom practices. If teaching performance tests are used to compare groups of teachers as Popham (1974) reported, then, of course, each teacher need attempt only one or two lessons.

Additional observations can be secured by using more students in the class and having each learner answer more test items. Even if a small class (e. g., six learners) and a short test (e. g., five items) are employed, there still would be a respectable number of observations ($6 \times 5 = 30$). Thus, from the viewpoint of reducing measurement error, it would seem much more profitable to increase the number of minilesson trials to the maximum number feasible.

3. Other Administrative Facets. There are several administrative factors which can be varied when teaching performance tests are used. Some of these include: (a) length of time the teacher is allowed to prepare for the instruction, (b) the time allowed for instruction, (c) the size of the class to be taught, and (d) whether or not a practice trial is permitted. Preliminary research (see, e. g., Millman 1974) suggests that these factors have but slight influence on the subsequent performance displayed by the learners.

Concluding Remarks

I and other advocates of teaching performance tests do not believe that these devices, as described this afternoon, should be the sole method for measuring teaching effectiveness. The ways in which good teaching can be evidenced are too numerous for any single criterion to tell the whole story. Nathan Gage (1968) in a research context wrote:

If it were necessary to sum it up in one word, my word would be analysis, breaking down the complexities that have proven to be so unmanageable when dealt with as a whole. We are no longer crippled by the notion that because there is one word "teaching," there is one, single, overall criterion of effectiveness in teaching that will take essentially the same form wherever teaching occurs.... It may well be that a 15-minute explanation of a five-page magazine article is still too large a unit of teaching behavior to yield valid, lawful knowledge. It may well be that the mean score on a 10-item test of comprehension, adjusted for student ability... is still too large and complex a dependent variable. But, compared with the massive, tangled, and unanalyzable units that have typically been studied in the past... such units seem precise and manageable indeed. (p. 606)

The cards do seem stacked against finding that single, valid measure of teaching effectiveness. We can now see that our earliest attempts at measuring a teacher's ability to bring about prespecified changes in learners by applying only a few teaching performance tests constructed on an opportunistic basis offered as much chance for success as completing an inside straight in draw poker. We should make assessments separately for those teaching situations considered most important or, if preferred, make a composite assessment containing samples of situations stratified by important task and administrative factors. The pot is too big, and our ante too small, to fold up our cards and quit.

References

Gage, Nathan L. An analytical approach to research on instructional methods.
Phi Delta Kappan, 1968, 49, 601-606.

Gagné, Robert M. Educational technology and the learning process.
Educational Researcher, 1974, 3, 3-8.

Humphreys, Lloyd G. The organization of human abilities. American Psychologist, 1962, 17, 475-483.

Millman, Jason. Psychometric characteristics of performance tests of teaching effectiveness. Paper presented at the annual meeting of the American Educational Research Association, February 1973. Pp. 11.

Popham, W. James. Evaluating teacher education candidates with teaching performance tests. Paper presented at the annual meeting of the American Educational Research Association, April 1974, Pp. 8.